

(considerably) Reducing Annotation Need in Self-Explanatory Models for Lung Nodule Diagnosis

(cRedAnno 

J. Lu^{1,2}, C. Yin^{1,3}, O. Krause¹, K. Erleben¹, M. B. Nielsen², S. Darkner¹

¹ Department of Computer Science, University of Copenhagen, Denmark

² Department of Diagnostic Radiology, Rigshospitalet, Copenhagen University Hospital, Denmark

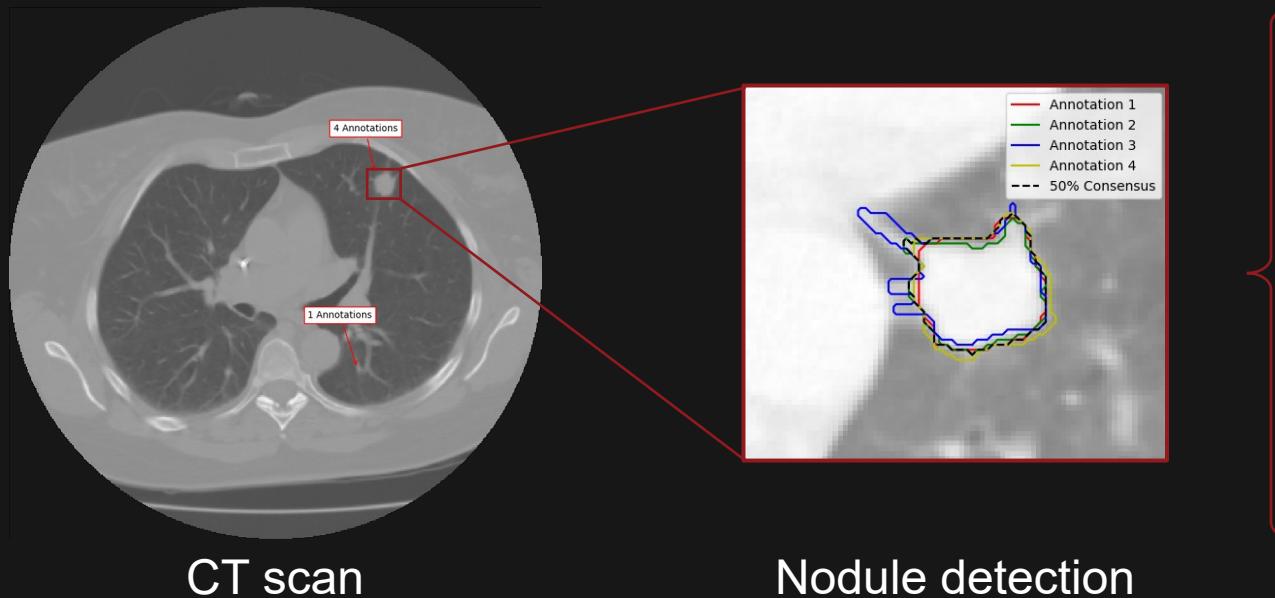
³ Department of Computer Science, Hong Kong Baptist University, China



KØBENHAVNS UNIVERSITET

E A

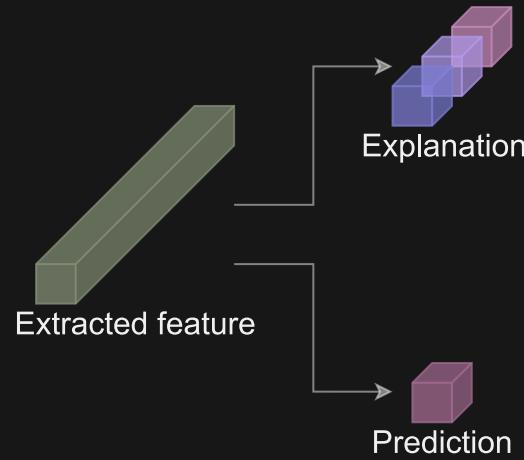
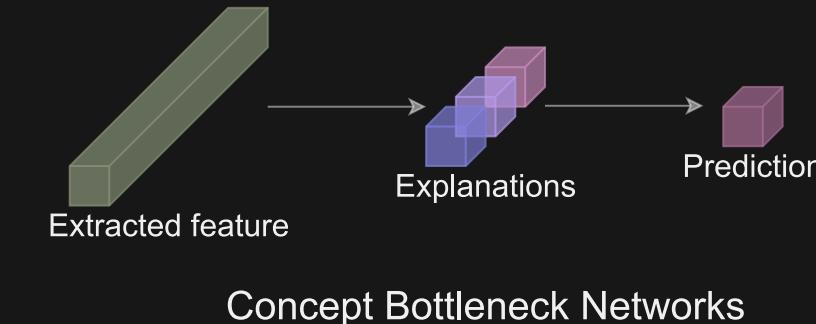
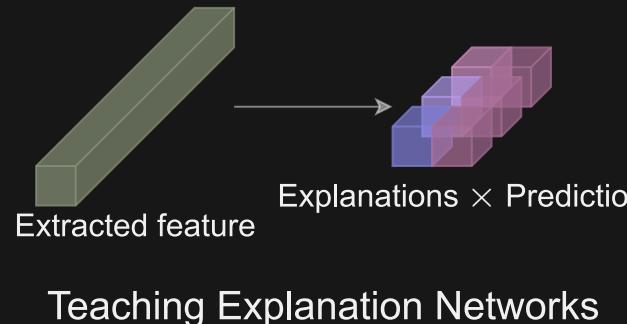
Background: Lung nodule diagnosis



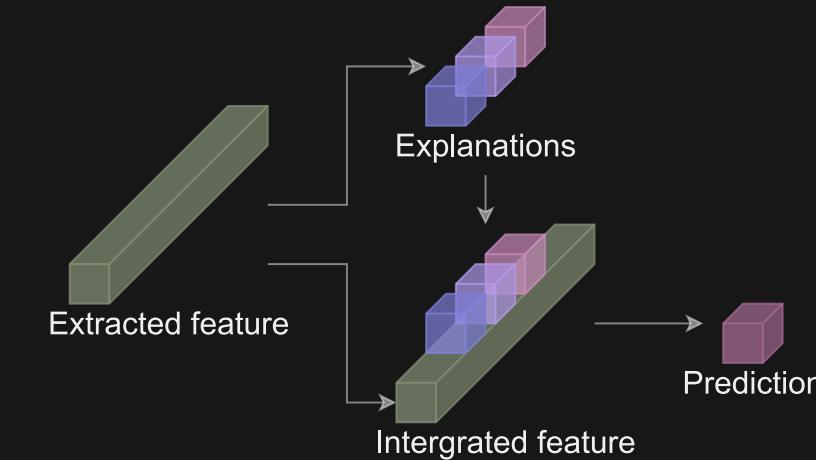
Nodule attributes	
Subtlety	Obvious
InternalStructure	Soft Tissue
Calcification	Non-central
Sphericity	Ovoid
Margin	Sharp
Lobulation	Nearly No Lobulation
Spiculation	Medium Spiculation
Texture	Solid
Malignancy	
Moderately Suspicious	

Annotation info

Related work: Feature-based self-explanatory models

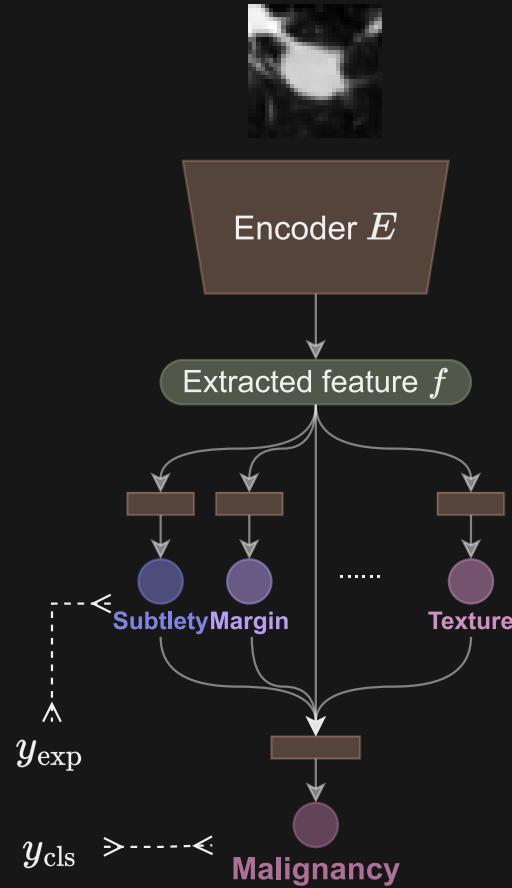


Multitask Learning Networks

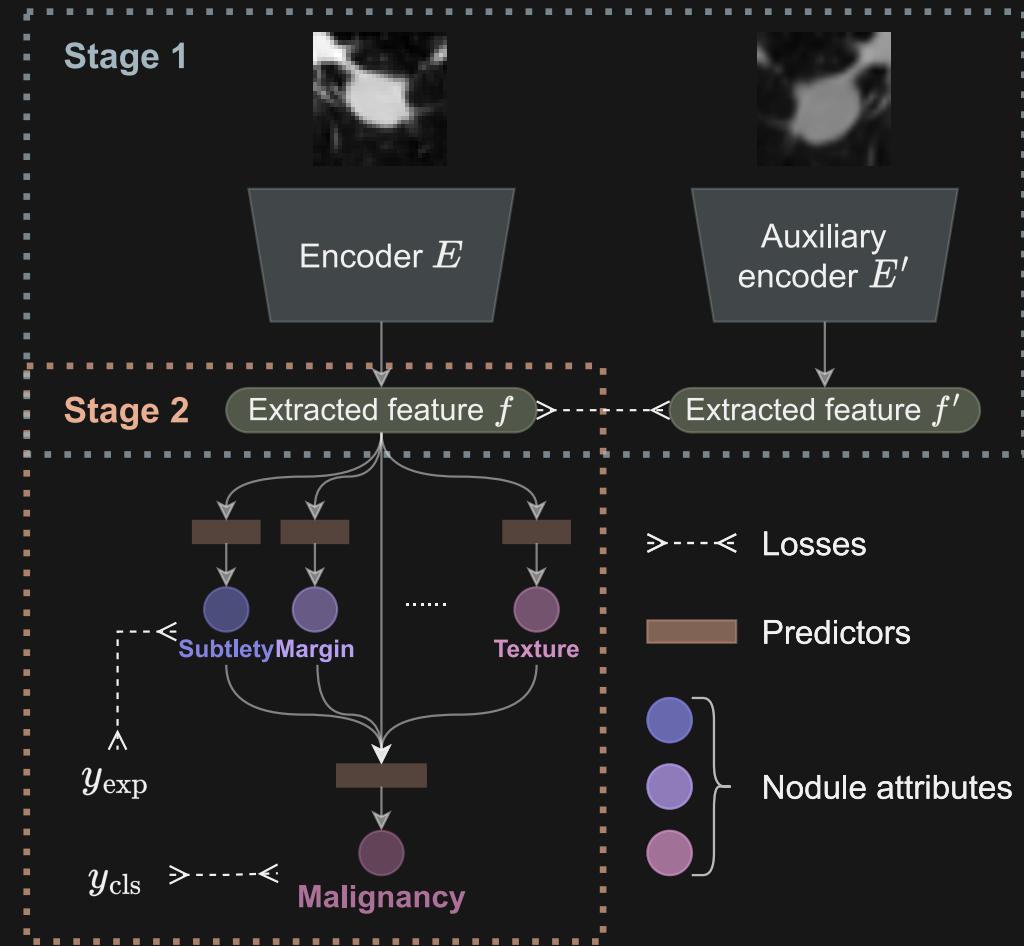


Hierarchical Networks

Method



Previous methods



cRedAnno

Results:

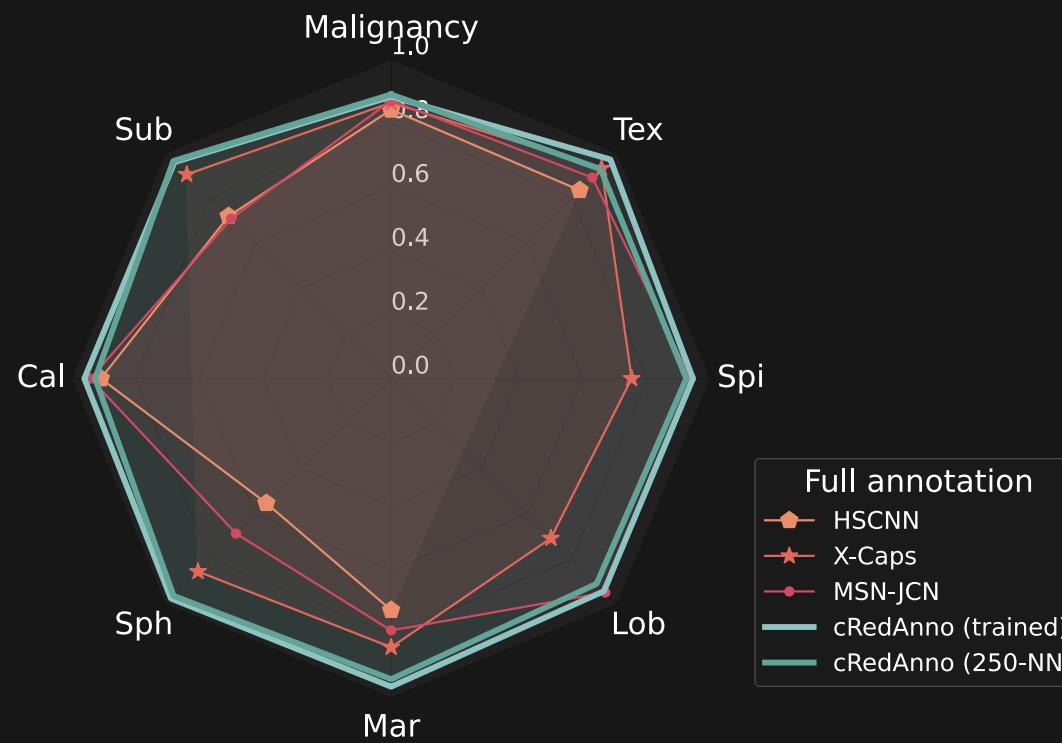
Predicting nodule attributes and malignancy

	Nodule attributes							Malignancy	#nODULES	Additional information																
	Sub	Cal	Sph	Mar	Lob	Spi	Tex																			
Full annotation																										
HSCNN ^[1]	71.90	90.80	55.20	72.50	-	-	83.40	84.20	4252	3D volume data																
X-Caps ^[2]	90.39	-	85.44	84.14	70.69	75.23	93.10	86.39	1149	None																
MSN-JCN ^[3]	70.77	94.07	68.63	78.88	94.75	93.75	89.00	87.07	2616	segmentation mask + diameter + OTSU + SLIC																
MTMR ^[4]	-	-	-	-	-	-	-	93.50	1422	all 2D slices in 3D volumes																
cRedAnno (50-NN)	94.93	92.72	95.58	93.76	91.29	92.72	94.67	87.52	730	<table border="1"> <thead> <tr> <th>#nODULES</th> <th>0</th> <th>1</th> <th>sum</th> </tr> </thead> <tbody> <tr> <td>train</td> <td>276</td> <td>242</td> <td>518</td> </tr> <tr> <td>val</td> <td>108</td> <td>104</td> <td>212</td> </tr> <tr> <td>sum</td> <td>384</td> <td>346</td> <td>730</td> </tr> </tbody> </table>	#nODULES	0	1	sum	train	276	242	518	val	108	104	212	sum	384	346	730
#nODULES	0	1	sum																							
train	276	242	518																							
val	108	104	212																							
sum	384	346	730																							
cRedAnno* (250-NN)	96.36	92.59	96.23	94.15	90.90	92.33	92.72	88.95																		
cRedAnno* (trained)	95.84	95.97	97.40	96.49	94.15	94.41	97.01	88.30																		
Partial annotation																										
WeakSup ^[5] (1:5)	43.10	63.90	42.40	58.50	40.60	38.70	51.20	82.40	2558	multi-scale 3D volume data, all malignancy annotations, 1/(1+N) attribute annotations																
WeakSup ^[5] (1:3)	66.80	91.50	66.40	79.60	74.30	81.40	82.20	89.10																		
cRedAnno (10%, 50-NN)	94.93	92.07	96.75	94.28	92.59	91.16	94.15	87.13																		
cRedAnno* (10%, 150-NN)	95.32	89.47	97.01	93.89	91.81	90.51	92.85	88.17	730	None																
cRedAnno* (1%, trained)	91.81	93.37	96.49	90.77	89.73	92.33	93.76	86.09																		

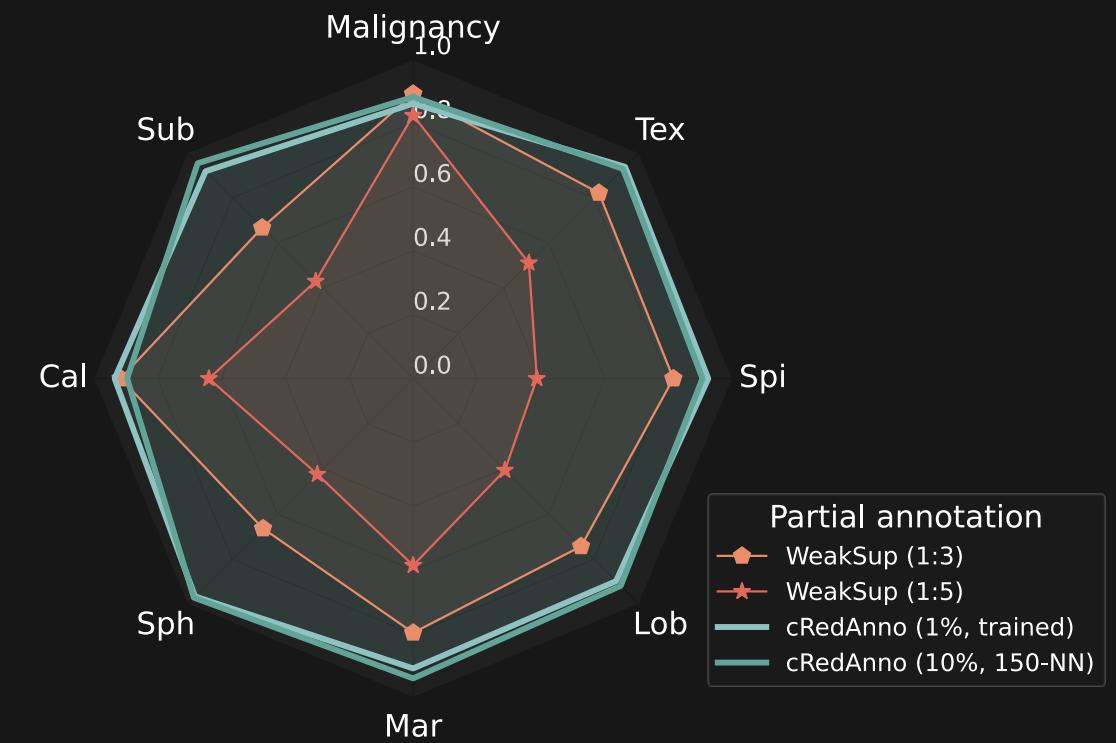
- [1] S. Shen *et al.*, "An interpretable deep hierarchical semantic convolutional neural network for lung nodule malignancy classification," *Expert Systems with Applications*, vol. 128, pp. 84–95, Aug. 2019.
- [2] R. LaLonde *et al.*, "Encoding Visual Attributes in Capsules for Explainable Medical Diagnoses," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020*, Cham, 2020, pp. 294–304.
- [3] W. Chen *et al.*, "End-to-End Multi-Task Learning for Lung Nodule Segmentation and Diagnosis," in *2020 25th International Conference on Pattern Recognition (ICPR)*, Milan, Italy, 2021, pp. 6710–6717.
- [4] L. Liu *et al.*, "Multi-Task Deep Model With Margin Ranking Loss for Lung Nodule Analysis," *IEEE Trans. Med. Imaging*, vol. 39, no. 3, pp. 718–728, Mar. 2020.
- [5] A. Joshi *et al.*, "Lung nodule malignancy classification with weakly supervised explanation generation," *J. Med. Imag.*, vol. 8, no. 04, Aug. 2021.

Results: Predicting nodule attributes and malignancy

Full annotation

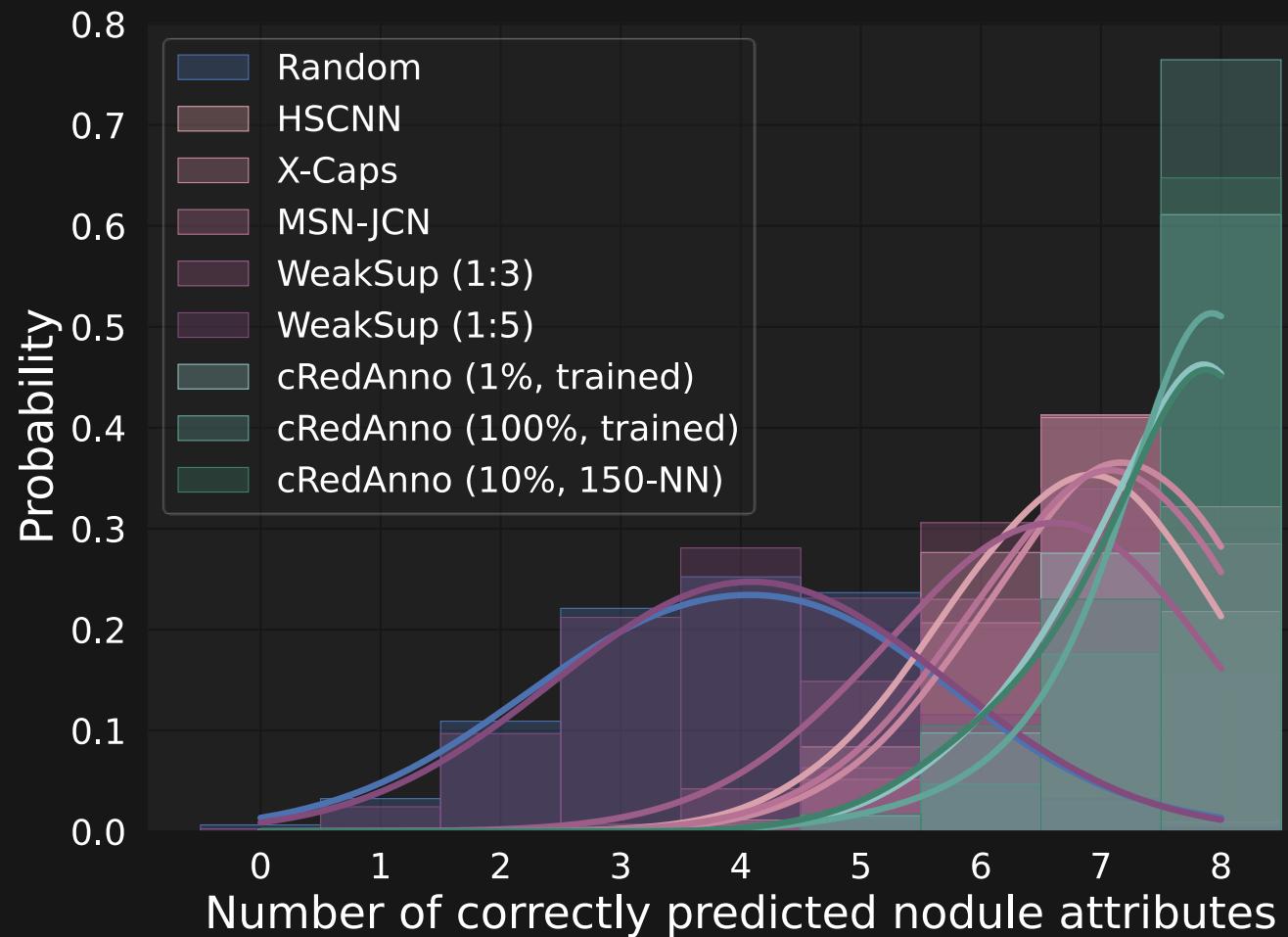


Partial annotation



Simultaneously high accuracy in predicting malignancy and all nodule attributes.

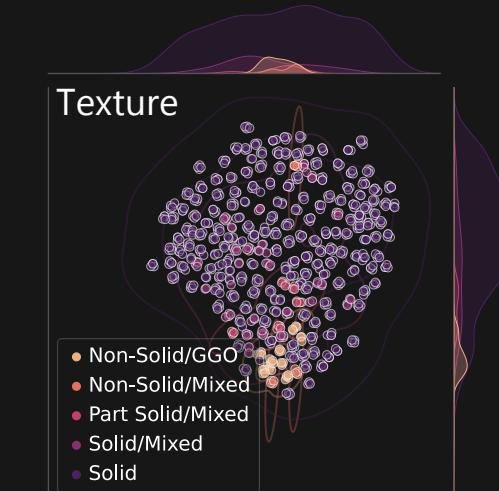
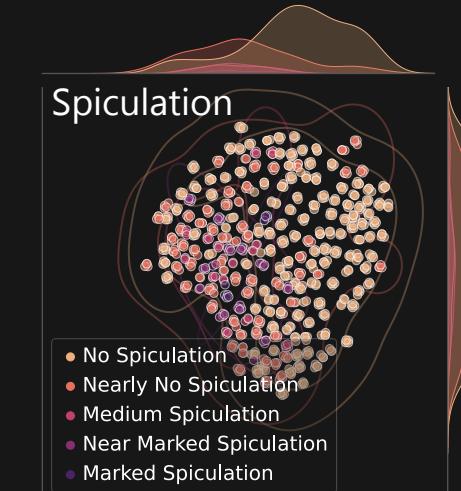
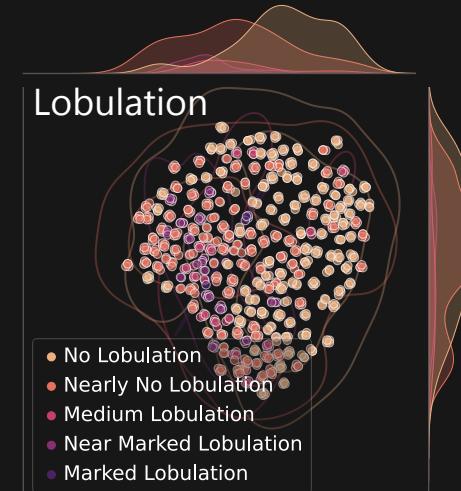
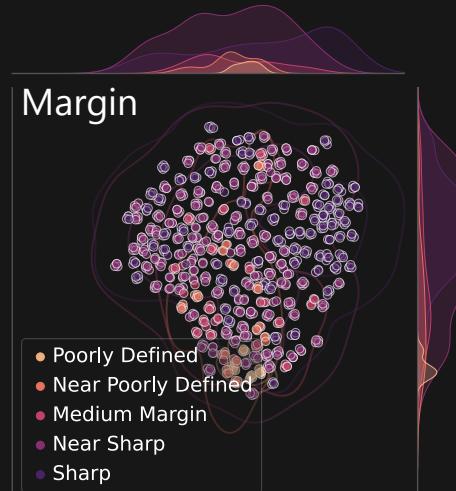
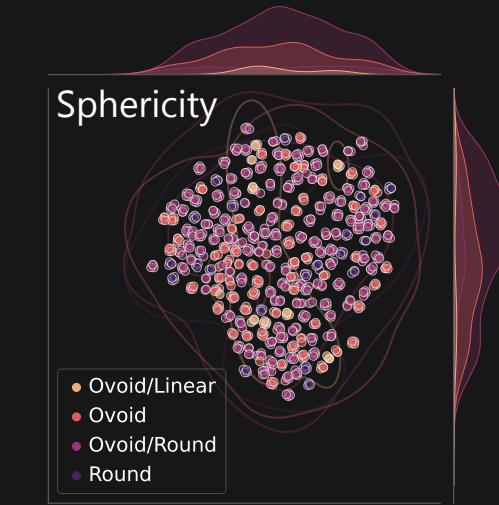
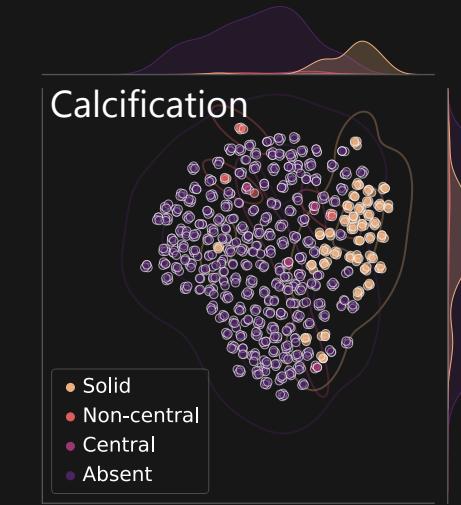
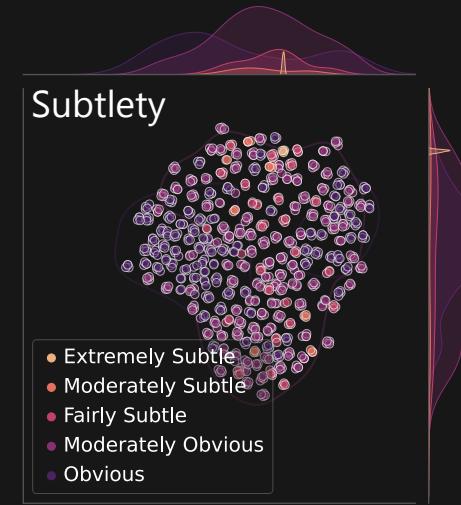
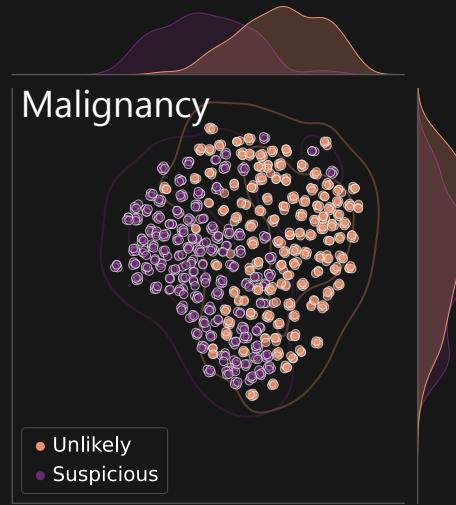
Results: Predicting nodule attributes



- cRedAnno shows a significantly larger probability of simultaneously predicting all 8 nodule attributes correctly.
- Approximately 90% nodules have at least 7 attributes correctly predicted.

Results:

Analysis of extracted features in learned space



Results:

Ablation study

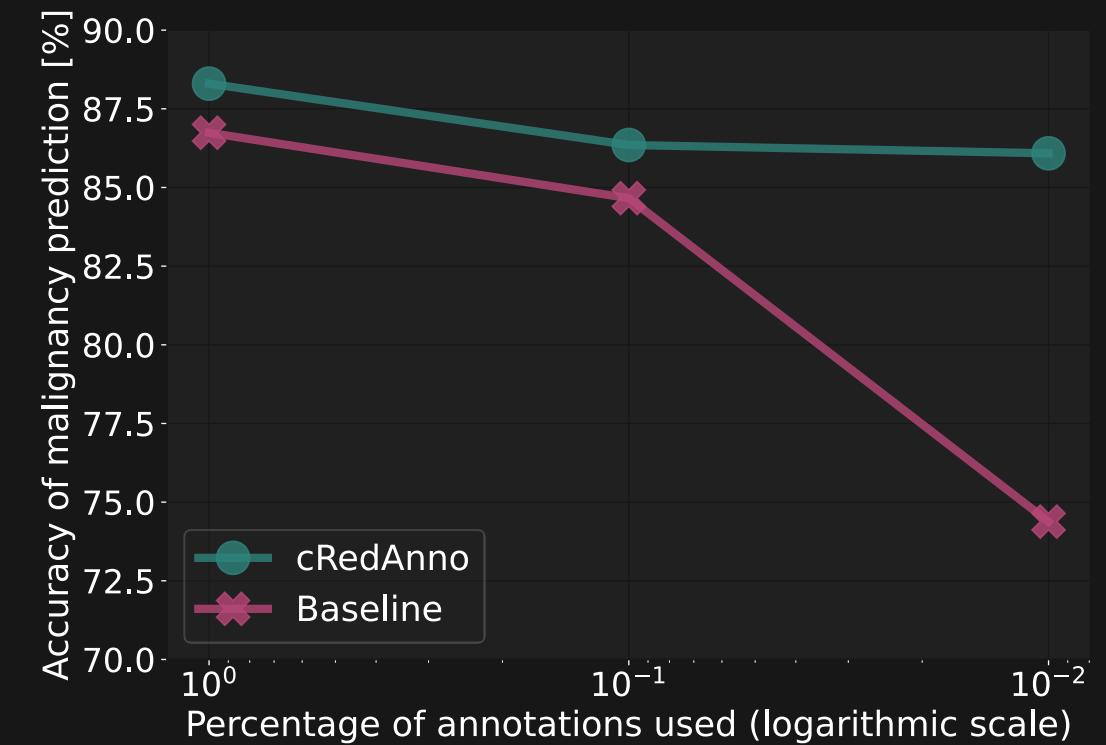
Validation of components

Accuracy of malignancy prediction (%). All annotations are used during training.

Arch	#params	Training strategy	ImageNet pretrain	Acc
ResNet-50	23.5M	end-to-end	✗	86.74*
		two-stage	✗	70.48
		two-stage	✓	70.48
ViT	21.7M	end-to-end	✗	64.24
		two-stage	✗	79.19
		two-stage	✓	88.30

* Representative setting and performance of previous works using CNN architecture

Annotation reduction



Conclusion

- A data-/annotation-efficient self-explanatory approach for lung nodule diagnosis
- Comparing with SOTA:
 - 1% annotation, fewer samples
 - comparable in malignancy prediction
 - significantly better in predicting all nodule attributes as explanations
- Visualising the learned space:
 - extracted features are highly separable
 - clustering coincides with clinical knowledge

- Open-source code



github.com/diku-dk/credanno

- Implementation
- Sample selection
- Pre-processing
- Experiments
- Plots

Considerably Reducing Annotation Need in Self-Explanatory Models for Lung Nodule Diagnosis (cRedAnno



github.com/diku-dk/credanno

CT scan

Nodule detection

Method

Nodule attributes	
Subtlety	Obvious
InternalStructure	Soft Tissue
Calcification	Non-central
Sphericity	Ovoid
Margin	Sharp
Lobulation	Nearly No Lobulation
Spiculation	Medium Spiculation
Texture	Solid
Malignancy	Moderately Suspicious

Annotation info

Method

Previous methods

The diagram shows a single-stage architecture. An input image is processed by an **Encoder E** to extract features f . These features are then processed by three parallel predictors (Subtlety, Margin, and Texture) to produce node attributes (SubtletyMargin and Texture). These attributes are combined to form a final prediction y_{cls} and a confidence measure y_{exp} .

cRedAnno

This diagram illustrates a two-stage architecture:

- Stage 1:** An input image is processed by an **Encoder E** to extract features f .
- Stage 2:** The features f from Stage 1 are combined with features f' from an auxiliary encoder E' . These combined features are then processed by three parallel predictors (Subtlety, Margin, and Texture) to produce node attributes (SubtletyMargin and Texture). These attributes are combined to form a final prediction y_{cls} and a confidence measure y_{exp} .

Legend:

- >--< Losses
- Predictors
- { Nodule attributes

Results

The figure displays results for nine features: Malignancy, Subtlety, Calcification, Sphericity, Margin, Lobulation, Spiculation, and Texture. Each feature is represented by a radar chart and a corresponding scatter plot.

- Malignancy:** Radar chart shows scores for Sub (0.5), Tex (0.5), Spi (0.1), Lob (0.8), Sph (0.1), Mar (0.1), and Cal (0.9). Scatter plot shows points colored by annotation type: HSCNN (orange), X-Caps (red), MSN-JCN (pink), cRedAnno (trained) (teal), and cRedAnno (250-NN) (dark teal).
- Subtlety:** Radar chart shows scores for Sub (0.5), Tex (0.5), Spi (0.1), Lob (0.8), Sph (0.1), Mar (0.1), and Cal (0.9). Scatter plot shows points colored by subtlety level: Extremely Subtle (orange), Moderately Subtle (red), Fairly Subtle (pink), Moderately Obvious (purple), and Obvious (dark purple).
- Calcification:** Radar chart shows scores for Sub (0.5), Tex (0.5), Spi (0.1), Lob (0.8), Sph (0.1), Mar (0.1), and Cal (0.9). Scatter plot shows points colored by calcification type: Solid (black), Non-central (red), Central (pink), and Absent (white).
- Sphericity:** Radar chart shows scores for Sub (0.5), Tex (0.5), Spi (0.1), Lob (0.8), Sph (0.1), Mar (0.1), and Cal (0.9). Scatter plot shows points colored by sphericity type: Ovoid/Linear (orange), Ovoid (red), Oval/Round (pink), and Round (dark purple).
- Margin:** Radar chart shows scores for Sub (0.5), Tex (0.5), Spi (0.1), Lob (0.8), Sph (0.1), Mar (0.1), and Cal (0.9). Scatter plot shows points colored by margin type: Poorly Defined (orange), Near Poorly Defined (red), Medium Margin (pink), Near Sharp (purple), and Sharp (dark purple).
- Lobulation:** Radar chart shows scores for Sub (0.5), Tex (0.5), Spi (0.1), Lob (0.8), Sph (0.1), Mar (0.1), and Cal (0.9). Scatter plot shows points colored by lobulation type: No Lobulation (orange), Nearly No Lobulation (red), Medium Lobulation (pink), Near Marked Lobulation (purple), and Marked Lobulation (dark purple).
- Spiculation:** Radar chart shows scores for Sub (0.5), Tex (0.5), Spi (0.1), Lob (0.8), Sph (0.1), Mar (0.1), and Cal (0.9). Scatter plot shows points colored by spiculation type: No Spiculation (orange), Nearly No Spiculation (red), Medium Spiculation (pink), Near Marked Spiculation (purple), and Marked Spiculation (dark purple).
- Texture:** Radar chart shows scores for Sub (0.5), Tex (0.5), Spi (0.1), Lob (0.8), Sph (0.1), Mar (0.1), and Cal (0.9). Scatter plot shows points colored by texture type: Non-Solid/GGO (orange), Non-Solid/Mixed (red), Part Solid/Mixed (pink), Solid/Mixed (purple), and Solid (dark purple).